HAL
open science

Time-aware Egocentric network-based User Profiling

Marie-Françoise Canut, Sirinya On-At, André Péninou, Florence Sèdes

To cite this version:

Marie-Françoise Canut, Sirinya On-At, André Péninou, Florence Sèdes. Time-aware Egocentric network-based User Profiling. IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM 2015), Aug 2015, Paris, France. pp.569-572, 10.1145/2808797.2809415 . hal-01343052

HAL Id: hal-01343052

https://hal.science/hal-01343052v1

Submitted on 7 Jul 2016

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Open Archive TOULOUSE Archive Ouverte (OATAO)
OATAO is an open access repository that collects the work of Toulouse researchers and makes it freely available over the web where possible.

This is an author-deposited version published in : http://oatao.univ-toulouse.fr/
Eprints ID : 15356

The contribution was presented at ASONAM 2015 :
http://asonam.cpsc.ucalgary.ca/2015/

To cite this version : Canut, Marie-Françoise and On-At, Sirinya and Péninou, André and Sèdes, Florence *Time-aware Egocentric network-based User Profiling*. (2015) In: 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM 2015), 25 August 2015 - 28 August 2015 (Paris, France).

Any correspondence concerning this service should be sent to the repository administrator: staff-oatao@listes-diff.inp-toulouse.fr

Time-aware Egocentric network-based User Profiling

Marie-Françoise Canut, Sirinya On-At, André Péninou and Florence Sèdes
IRIT, University of Toulouse, UMR CNRS 5505, 31062 TOULOUSE Cedex 9
{marie-francoise.canut, sirinya.on-at, andre.peninou, florence.sedes}@irit.fr

***Abstract*— Improving the egocentric network-based user's profile building process by taking into account the dynamic characteristics of social networks can be relevant in many applications. To achieve this aim, we propose to apply a time-aware method into an existing egocentric-based user profiling process, based on previous contributions of our team. The aim of this strategy is to weight user's interests according to their relevance and freshness. The time awareness weight of an interest is computed by combining the relevance of individuals in the user's egocentric network (computed by taking into account the freshness of their ties) with the information relevance (computed by taking into account its freshness). The experiments on scientific publications networks (DBLP/Mendeley) allow us to demonstrate the effectiveness of our proposition compared to the existing time-agnostic egocentric network-based user profiling process.**

Keywords*— *User's Profile, Social Network, Egocentric Network, Time-aware Method

I. INTRODUCTION

User's profile is a key element to improve adaptive information mechanisms (e.g. recommendation, personalization, customization). These mechanisms rely on user's profile that collects and identifies user personal information and interests in order to propose adapted information with regards to the specific needs of the user. In our context of work, we consider the user's profile as a vector of weighted user's interests represented by keywords. User's profile can be built either by using the user's behavior or by using external sources [1].

We are interested here in building user's profile using the information from the members of the user's social network. For this, we have to extract user's interests from the relations and interactions of the user's direct social contacts called "egocentric network", that is to say people directly connected to him/her (the user). We call "social profile" the user's profile built within this approach. The social profile can be used as an additional profile with regard to the (classical) profile. Such an extra profile can be useful for instance with inactive users, whose profiles are empty or very poor.

The Online Social Networks services enable users to contribute, broadcast online contents and connect with others who share the same interests and/or activities. So, users' online behavior becomes more and more evolving. Extracting user's interests from this kind of information with an evolution-agnostic technique may lead to misinterpretation of his/her long-term accurate interests.

In this work, we focus on taking into account the evolution of user's interests in social network-based user profiling process in order to build a more relevant and up-to-date social profile. We try to answer the following problems: (i) how to select the relevant individuals in the user social network as meaningful information sources? (ii) how to select relevant information from these selected individuals to extract the relevant and fresh interests to build the user social profile? To answer these two issues, we propose to apply a time-aware method to social profiling process, in order to weight each extracted interest according to its relevance and freshness. To evaluate our approach, we apply a time-aware function to the community-based user profiling process, previously proposed in our team by Tchuente [2] in order to study the impact of a time-aware approach compared to a time-agnostic one.

The paper is structured as follows. In the next section, the existing works on user profiling from egocentric network are synthesized. In section three, we present our study on techniques proposed to deal with user's interests evolution. In section four, we define the experimentation process. The last section concludes with some future works.

II. INITIAL CONTRIBUTION: EGOCENTRIC NETWORK-BASED USER PROFILING

In the literature, several techniques have been proposed to extract user's interests and build or enrich the user's profile [1]. Among all, one effective technique consists in using information from user's social network [3]; such an approach is called social network-based user profiling. In this approach, user's interests are extracted from user's social network information. We call "social profile" the user's profile built within this approach. Social profile is useful, on the one hand, to provide additional information that can improve the relevance of existing user's profile and, on the other hand, to overcome the cold start problem or to complete non-existing/missing profile of less active users.

In this context, our team contribution [2] proposes a community-based algorithm to derive user's social profile. This work considers that the user is better described by communities of people around him, especially people that are directly connected to him (his/her egocentric network). We consider in this work, the user's egocentric network as a specific social network that takes into account only his/her direct neighbors (direct relations). We describe the egocentric network as follows: for each user u, we consider the undirected graph G(u) = (V, E) where V is the set of

nodes (individuals) directly connected to u, and E is the set of relationships between each pair of nodes of V. In [2], the user's profile is composed of two dimensions. The first one contains the user's interests computed from the sole user's behavior. The second one contains the user's interests computed by using the behavior of people involved in the user's egocentric network. This work introduces a user's social dimension building process named CoBSP (Community Based Social Profile), consisting of four steps: community extraction, community profiling, interest weight calculation and social dimension derivation. The detail of this process can be found in [2].

The performance of CoBSP algorithm has been proved with empirical results. However, the evolution of the user's interests has not been taken into consideration in this existing work whereas user online behavior tends to change over time. We are interested in taking into account this characteristic in order to improve the effectiveness of social network-based user profiling approach.

III. USER'S INTERESTS EVOLUTION DEALING TECHNIQUE

To deal with user's interests evolution, it is essential to take into account the time factor. Several works propose time-aware approaches to cope with the evolution of their studied data. Some of these works rely on the time window or time forgetting approach that selects only the information from the latest periods [4]. Within this approach, outdated information is completely forgotten. However, in some cases, the ignored information could eventually be valuable. Thus, this can lead to the lost of useful knowledge. This might be reasonable when the time shift is abrupt, but less so when time shift is gradual. Some other works propose to use long-term user's profile [5], in which the whole user's information is retained to extract his/her interests. However, this approach could not always reflect relevant user's interests as it gives the same importance to all information whereas some information could be out-of-date. To avoid this drawback, some works propose a time decay approach. This approach proposes to assign the greater value to recent information [6]. For this, the exponential time function is widely used to weight information by gradually decreasing the weight of the older one [7].

In our work context, information sharing from social networks relies on each user's behavior. Some users are very active and frequently share information. Some users are less active. To apply a time-aware approach that fits all user types, the time decay approach is more appropriate for our work. We are interested in extracting user's interests from his/her social network to build his/her social profile. To build an effective user social profile, it is necessary to take into account the evolution characteristics of Online Social Networks in term of network structure and information flow [8], [9]. The former provides volatile characteristic of relations between users. The latter provides volatile characteristic of shared information that reflects the evolution of user's preferences and interests. Contrary to the existing works, to deal with the interest evolution, we apply a time criteria according to these two factors.

IV. PROPOSED METHOD

We consider in this paper, the user's profile model proposed by [2] and focus in particular on improving the social dimension derivation process. We propose to modify the user's interests extraction step of the existing CoBSP. In this step, we apply a time-aware function in order to weight each interest according to its relevance (called structuro-semantico-temporal weight: StrSemTem). This weight is computed by combining: i) the structural time-aware weight (called structuro-temporal weight: StrTem), computed by applying a time-aware link prediction method, ii) the time-aware weight of the information used to extract this interest (called semantico-temporal weight: SemTem), computed by applying a term frequency method and a time-weighted function.

A. Time-aware method calculation

1) Structuro-temporal weight (StrTem)

This weight consists in computing the relevance of relations between the user and each member in his egocentric network by taking into account a time criteria. We propose exploiting temporal link prediction technique, which is widely used in graph mining and network analysis, as a relevant measure between the user and each member in his egocentric network. It consists in predicting, from a given network, which pairs of nodes are likely to link together [10]. The link prediction has a wide variety of application such as detecting new relations in social networks, identifying the collaboration in organizations. Unlike its initial use, we apply this technique in order to compute a similarity score between two connected nodes to get their relationship strength that could be exploited to predict their relation in the future. In our work context, we consider this similarity score as the relevant weight between these two nodes. We adopt the work of [11] that proposes to apply a temporal criterion (age of the last interaction between two studied nodes) on an existing link prediction method like Adamic/Adar, a method based on the common neighbors metric, introduced in [10]. The similarity score is calculated by the following formula (1):

$$AA_{tmp_{score(x,y)}} = \sum_{z \epsilon \Gamma x \cap \Gamma y} \frac{w(x,z) \cdot w(z,y)}{log|\Gamma(z)|} \quad (1)$$

where w(x,y) is the temporal weight between the nodes x and y (according to their latest interaction timestamp). In this paper, to compute this temporal weight, we adopt a time-exponential function [7] that weights a given item according to its publication date.

$$f(t) = e^{-\lambda t} \quad (2)$$

For each t_i ($i \in N$), we consider t_0 as the most recent moment (ex: t_0 for 2014, t_1 for 2013 ...). The parameter represents the decay rate. The higher the value of λ is, the faster old data decays and the lower is the importance of the historical information compared to more recent data. The value of λ will be fixed in the experiments. To obtain the time-aware structural weight, we multiply the calculated time-aware Adamic/Adar score of each individual with the structural score (centrality degree) of his/her community (see the CoBSP process). Thus, for each individual **ind** from the

community **c** of the egocentric network of a user **u**, his time-aware structural weight is computed as follows:

$$StrTem(ind) = AA_tmp_score(u, ind) \quad (3)$$

2) Semantico-temporal weight (SemTem)

To compute this weight, we first apply temporal weight to each information according its freshness using the same time-exponential function (2). The more the information is recent, the more it is considered as important and the higher its weight is. Then we extract the keywords from this information and compute their term frequency weight (tf). Thus, the semantic-temporal weight is calculated according to formula (4).

$$semTem(i) = score_tf(i) * temporal_weight(i) \quad (4)$$

3) Final weight (structuro-semantico-temporal weight)

The final temporal weight of an interest **i** extracted from an individual **ind** is computed by combining its structuro-temporal and its semantico-temporal weights varying with a parameter γ that will be fixed in the experiments:

$$StrSemTem(i,ind) = \gamma(StrTem(ind)) + (1-\gamma)SemTem(i) \quad (5)$$

V. EXPERIMENTS

We consider in this experiment the notions of user dimension and social dimension to distinguish between the user's own interests (user dimension) and the interests extracted from his egocentric network (social dimension). The evaluation consists in comparing the proposed time-aware approach with one of the time-agnostic existing work (CoBSP [2]). The strategy of evaluation consists in looking between the existing CoBSP and our time-aware proposition, the one that builds the closest social dimension to the real profile of the user (user dimension).

To understand information dynamics awareness influence, we propose to distinguish two different algorithms from our proposition method. We call CoBSP_StrTem, the community-based algorithm that apply only the StrTem (formula (3)) and CoBSP_StrSemTem, the community-based algorithm that apply the combined weight StrSemTem (formula (5)). We have conducted the experiments on two co-authorship networks (Mendeley and DBLP). In order to avoid using identical data sources in the user and social dimensions, we have decided to adopt two distinct data sources to build the two dimensions: DBLP for the social dimension and Mendeley for the user dimension. We take into consideration in our experiment, the authors who exist in both DBLP and Mendeley and have a sufficient number of interests indicated in their Mendeley profile so that we can compare the results correctly. Subsequently, we have decided to take into account in our evaluation, the authors that have at least 50 co-authors and have at least six interests explicitly indicated in their Mendeley profile.

A. Profiles building process

We adopted in this work, the methodology process of building and evaluating authors' profiles from DBLP and Mendeley presented in [2].

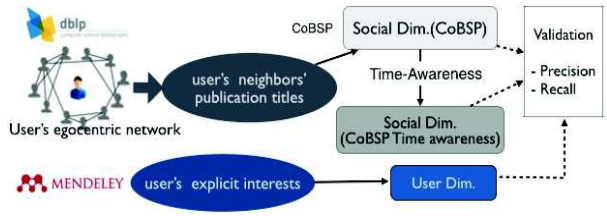

Fig. 1. Evaluation process

1) User dimension construction

This user dimension is built by mining keywords in the list of interests that the author explicitly indicates in his/her Mendeley profile. The first step consists in collecting the keywords in user's Mendeley profile. In the second step, we extract interests from the collected keywords using a text-mining engine: we first use dictionaries/thesaurus to merge the keywords having the same meaning. Then we apply filters to remove empty words, in order to retain only consistent interests. In the third step, we compute interests weights using a semantic weight. The semantic weight of each interest is computed by its tf measure that represents the term frequency of each interest in the set of all found interests.

2) Social dimension construction

For the existing social dimension building process (CoBSP) [2], interests are detected by mining texts that appear in the title of publications of communities or individuals according to the algorithm used to derive social dimension (CoBSP). To build social dimension of the proposed time-aware method, we use the same social dimension building process as previously presented. Except in the community profiling step (step 2 of the CoBSP), instead of using the tf measure, we give a weight to each interest extracted from users' social publication by using its structuro-temporal (StrTem) (formula 3) and its structuro-semantico-temporal (StrSemTem) (formula 5) weight to build successively the social dimension for the CoBSP_StrTemand CoBSP_StrSemTem algorithms.

B. Evaluation

To evaluate the relevance of each social dimension, we use the precision and the recall measures. The precision represents the proportion of relevant found items among the total number of items. In our experimentation context, the precision formula is presented as follows (6):

$$precision = \frac{Nb.Interests(Social\ Dimension \cap User\ Dimension)}{Nb.Interests(Social\ Dimension)} \quad (6)$$

The recall represents the proportion of relevant found items compared to the total number of relevant items .

$$recall = \frac{Nb.Interests\ (Social\ Dimension \cap User\ Dimension)}{Nb.Interests(User\ dimension)} \quad (7)$$

To compute the precision and the recall, we only consider the most relevant interests in the social dimension. Their number is top(N+m) where N is the total number of interests in the user dimension and m is a constant used to enlarge the targeted found interests (m=5 in this experiment).

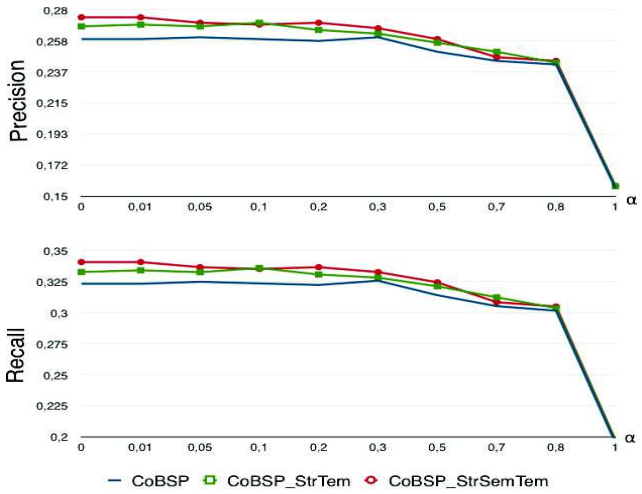

Fig. 2. Comparison of the precision (top) and the recall (bottom) of calculated social dimension of the COBSP algorithm (Blue), CoBSP_StrTem algorithm (Green) and CoBSP_StrSemTem algorithm (Red)

C. Results

The figure 2 showed above presents comparative curves (precision, recall) of the social dimensions for all users studied in this work. The alpha value indicates the weight of structural score compared to the semantic score as presented in [2]. Based on several experimentations using different values of γ, we compute temporal weight in the formula (5) with $\gamma = 0{,}75$. In the same way, we fixed the value of λ to 0.2.

We can see that CoBSP_StrTem algorithm outperforms the CoBSP one in term of precision and recall. The best result can be observed when $\alpha = 0.1$ with successively 4.3 % and 3.8 % precision and recall gain rate compared to CoBSP algorithm. The positive average gain values demonstrate the benefits of using temporal link prediction method during user social dimension building step. After applying the CoBSP_StrSemTem, which uses the combined temporal weight of the structuro-temporal weight and the semantico-temporal weight, the algorithm produces better results compared to the CoBSP_StrTem algorithm. It provides consequently a larger gap compared to the CoBSP algorithm, especially when $\alpha \in [0 ; 0{,}01]$, with successively 5.8 % and 5.4% gain rate in terms of precision and recall compared to CoBSP and 1.9 % and 2.0% compared to CoBSP_StrTem algorithm.

These results demonstrate the benefit of taking into account the temporal characteristic of the information to extract the more significant user's interests to build a more relevant user social profile. This demonstrates also the fact that the information diffusion takes also an important role in user's interests evolution in online social network, as does the social network structure.

VI. Conclusion and Future Works

In this paper, we suggest taking into account the social network dynamics in user's social profile building process. Considering the individuals of the user social networks and the contained information, we propose to use temporal criteria (time weights) to weight both individuals and information in order to extract relevant and up-to-date user's interests. The experimentations show the effectiveness of this method compared to the existing time-agnostic one. This demonstrates the benefit of taking into account the evolution of user's interests (by considering the social network evolution) in the social network-based user profiling process.

In short term, we plan to apply this approach in other social network fields where the dynamic characteristics of the network is more important in order to evaluate our proposition in a large-scale data. We also envisage studying other link prediction algorithms and other time-weight functions to enhance the performance of our approach. Our long term perspective consists in proposing a generic platform that extracts the information and builds the user social profile according to the type and the specific characteristics of any underlying social network. Such a platform would be parameterized by the characteristics of the targeted social network.

References

[1] S. Gauch, M. Speretta, A. Chandramouli, and A. Micarelli, "User Profiles for Personalized Information Access," in *The Adaptive Web*, P. Brusilovsky, A. Kobsa, and W. Nejdl, Eds. Springer Berlin Heidelberg, 2007, pp. 54–89.

[2] D. Tchuente, M.-F. Canut, N. Jessel, A. Peninou, and F. Sèdes, "A community-based algorithm for deriving users' profiles from egocentrics networks: experiment on Facebook and DBLP," *Soc. Netw. Anal. Min.*, vol. 3, no. 3, pp. 667–683, Sep. 2013.

[3] I. Guy, N. Zwerdling, D. Carmel, I. Ronen, E. Uziel, S. Yogev, and S. Ofek-Koifman, "Personalized Recommendation of Social Software Items Based on Social Relations," in *Proceedings of the Third ACM Conference on Recommender Systems*, New York, NY, USA, 2009, pp. 53–60.

[4] M. A. Maloof and R. S. Michalski, "Selecting Examples for Partial Memory Learning," *Mach. Learn.*, vol. 41, no. 1, pp. 27–52, Oct. 2000.

[5] B. Tan, X. Shen, and C. Zhai, "Mining Long-term Search History to Improve Search Accuracy," in *Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, New York, NY, USA, 2006, pp. 718–723.

[6] N. Zheng and Q. Li, "A recommender system based on tag and time information for social tagging systems," *Expert Syst. Appl.*, vol. 38, no. 4, pp. 4575–4587, Apr. 2011.

[7] Y. Ding and X. Li, "Time Weight Collaborative Filtering," in *Proceedings of the 14th ACM International Conference on Information and Knowledge Management*, New York, NY, USA, 2005, pp. 485–492.

[8] M. T. Rivera, S. B. Soderstrom, and B. Uzzi, "Dynamics of Dyads in Social Networks: Assortative, Relational, and Proximity Mechanisms," *Annu. Rev. Sociol.*, vol. 36, no. 1, pp. 91–115, 2010.

[9] L. Weng, J. Ratkiewicz, N. Perra, B. Gonçalves, C. Castillo, F. Bonchi, R. Schifanella, F. Menczer, and A. Flammini, "The Role of Information Diffusion in the Evolution of Social Networks," in *Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, New York, NY, USA, 2013, pp. 356–364.

[10] D. Liben-Nowell and J. Kleinberg, "The Link Prediction Problem for Social Networks," in *Proceedings of the Twelfth International Conference on Information and Knowledge Management*, New York, NY, USA, 2003, pp. 556–559.

[11] T. Tylenda, R. Angelova, and S. Bedathur, "Towards Time-aware Link Prediction in Evolving Social Networks," in *Proceedings of the 3rd Workshop on Social Network Mining and Analysis*, New York, NY, USA, 2009, pp. 9:1–9:10.